CSE 442 - Data Visualization

Visual Encoding Design

Jeffrey Heer  University of Washington
A Design Space of Visual Encodings
Mapping Data to Visual Variables

Assign **data fields** (e.g., with N, O, Q types) to **visual channels** (x, y, color, shape, size, ...) for a chosen **graphical mark** type (point, bar, line, ...).

Additional concerns include choosing appropriate **encoding parameters** (log scale, sorting, ...) and **data transformations** (bin, group, aggregate, ...).

These options define a large combinatorial space, containing both useful and questionable charts!
Expressive?

Raw

Aggregate (Count)
1D: Quantitative

Raw

Aggregate (Count)
Expressive?

Raw

Aggregate (Count)
Effective?

Raw

Aggregate (Count)

![Bar chart showing data distribution and count for Miles_per_Gallon.](image)
Raw (with Layout Algorithm)

Treemap

Bubble Chart

Aggregate (Distributions)

interquartile range
(middle 50%)

Box Plot

Violin Plot
2D: Nominal x Nominal

Raw

Origin
Europe
Japan
USA

Cylinders

Aggregate (Count)

Origin
Europe
Japan
USA

Cylinders

COUNT

3
108

COUNT

20
40
80
100
2D: Quantitative x Quantitative

Raw

Aggregate (Count)
2D: Nominal x Quantitative

Raw

Aggregate (Mean)

Origin

Europe
Japan
USA

Miles_per_Gallon

Mean(Miles_per_Gallon)

Origin

Europe
Japan
USA

Miles_per_Gallon

Mean(Miles_per_Gallon)
Raw (with Layout Algorithm)

Treemap

Bubble Chart

Beeswarm Plot

Origin
- Europe
- Japan
- USA
3D and Higher

Two variables $[x, y]$  
Can map to 2D points.  
Scatterplots, maps, ...

Third variable $[z]$  
Often use one of size, color, opacity, shape, etc. Or, one can further partition space.  
What about 3D rendering?
Other Visual Encoding Channels?

Wind Map

April 1, 2015
11:35 pm EST
(time of forecast download)

Top speed: 30.5 mph
Average: 10.2 mph
Encoding Effectiveness
Effectiveness Rankings

QUANTITATIVE
Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL
Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL
Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume
<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area (Size)</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td>Density (Value)</td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td>Color Sat</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area (Size)</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>
Effectiveness Rankings

QUANTITATIVE
Position
Length
Angle
Slope
Area (Size)
Volume
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Shape

ORDINAL
Position
Density (Value)
Color Sat
Color Hue
Texture
Connection
Containment
Length
Angle
Slope
Area (Size)
Volume
Shape

NOMINAL
Position
Color Hue
Texture
Connection
Containment
Density (Value)
Color Sat
Shape
Length
Angle
Slope
Area
Volume

[Mackinlay 86]
Color Encoding (Choropleth Map)
# Effectiveness Rankings

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td><strong>Area (Size)</strong></td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td><strong>Density (Value)</strong></td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td>Color Sat</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Color Hue</td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area (Size)</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>
Color Encoding (Choropleth Map)
Area Encoding (Symbol Map)
Gene Expression Time-Series [Meyer et al. '11]

Color Encoding
Effectiveness Rankings

**QUANTITATIVE**
- Position
- Length
- Angle
- Slope
- Area (Size)
- Volume
- **Density (Value)**
- Color Sat
- **Color Hue**
- Texture
- Connection
- Containment
- Shape

**ORDINAL**
- Position
- Density (Value)
- Color Sat
- Color Hue
- Texture
- Connection
- Containment
- Volume
- Shape

**NOMINAL**
- Position
- Color Hue
- Texture
- Connection
- Containment
- Density (Value)
- Color Sat
- Shape
- Volume
- Area
Gene Expression Time-Series [Meyer et al ‘11]

Color Encoding

Position Encoding
Artery Visualization [Borkin et al. '11]

Rainbow Palette

2D

62%

Diverging Palette

2D

92%

3D

39%

71%
# Effectiveness Rankings

<table>
<thead>
<tr>
<th>QUANTITATIVE</th>
<th>ORDINAL</th>
<th>NOMINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
<td><strong>Position</strong></td>
</tr>
<tr>
<td>Length</td>
<td>Density (Value)</td>
<td>Color Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Color Sat</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Color Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area (Size)</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density (Value)</td>
</tr>
<tr>
<td><strong>Density (Value)</strong></td>
<td>Containment</td>
<td>Color Sat</td>
</tr>
<tr>
<td><strong>Color Sat</strong></td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td><strong>Color Hue</strong></td>
<td>Angle</td>
<td>Length</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Angle</td>
</tr>
<tr>
<td>Connection</td>
<td>Area (Size)</td>
<td>Slope</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Area</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Volume</td>
</tr>
</tbody>
</table>
Scales & Axes
Include Zero in Axis Scale?

Government payrolls in 1937 [How To Lie With Statistics. Huff]
Include Zero in Axis Scale?

Yearly CO$_2$ concentrations  [Cleveland 85]
Include Zero in Axis Scale?

Violates Expressiveness Principle!

Compare Proportions (Q-Ratio)

Compare Relative Position (Q-Interval)
Axis Tick Mark Selection

What are some properties of “good” tick marks?
**Simplicity** - numbers are multiples of 10, 5, 2

**Coverage** - ticks near the ends of the data

**Density** - not too many, nor too few

**Legibility** - whitespace, horizontal text, size
How to Scale the Axis?
One Option: Clip Outliers
Clearly Mark Scale Breaks

Violates Expressiveness Principle!

Poor scale break [Cleveland 85]  Well-marked scale break [Cleveland 85]
Scale Break vs. Log Scale

[Cleveland 85]
Scale Break vs. Log Scale

Both increase visual resolution
Scale break: difficult to compare (cognitive – not perceptual – work)
Log scale: direct comparison of all data
Logarithms turn *multiplication* into *addition*.

\[ \log(x \cdot y) = \log(x) + \log(y) \]

Equal steps on a log scale correspond to equal changes to a multiplicative scale factor.
Linear Scale vs. Log Scale

Linear Scale

Log Scale
Linear Scale vs. Log Scale

**Linear Scale**
Absolute change

**Log Scale**
Small fluctuations
Percent change
\[ d(10,30) > d(30,60) \]
Bending the Curve

Logarithmic scales can emphasize the rate of change in a way that linear scales do not. Italy seems to be slowing the coronavirus infection rate, while the number of cases in the United States continues to double every few days.

By The New York Times | Data from Worldometer
When To Apply a Log Scale?

Address data skew (e.g., long tails, outliers)
Enables comparison within and across multiple orders of magnitude.

Focus on multiplicative factors (not additive)
Recall that the logarithm transforms $\times$ to $+$!
Percentage change, not linear difference.

Constraint: positive, non-zero values
Constraint: audience familiarity?
Multidimensional Data
Visual Encoding Variables

- Position (X)
- Position (Y)
- Area
- Value
- Texture
- Color
- Orientation
- Shape

~8 dimensions?
Example: Coffee Sales

Sales figures for a fictional coffee chain

Sales | Q-Ratio
---|---
Profit | Q-Ratio
Marketing | Q-Ratio
Product Type | N {Coffee, Espresso, Herbal Tea, Tea}
Market | N {Central, East, South, West}
Encode “Sales” (Q) and “Profit” (Q) using Position.
Encode “Product Type” (N) using Hue
Encode “Market” (N) using Shape.
Encode “Marketing” \( (Q) \) using Size
A trellis plot subdivides space to enable comparison across multiple plots. Typically nominal or ordinal variables are used as dimensions for subdivision.
Small Multiples

[MacEachren ’95, Figure 2.11, p. 38]
Small Multiples

[MacEachren '95, Figure 2.11, p. 38]
Scatterplot Matrix (SPLOM)

Scatter plots for pairwise comparison of each data dimension.
Parallel Coordinates
Parallel Coordinates [Inselberg]
Parallel Coordinates [Inselberg]

Visualize up to ~two dozen dimensions at once
1. Draw parallel axes for each variable
2. For each tuple, connect points on each axis
Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.
Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.
Highly sensitive to axis **scale** and **ordering**.
Expertise required to use effectively!
Visual Encoding Design

Use **expressive** and **effective** encodings

**Reduce** the problem space

Avoid **over-encoding**

Use **space** and **small multiples** intelligently

Use **interaction** to generate **relevant** views

Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!
About the design process...

Visualization draws upon both science and art! Principles like expressiveness & effectiveness are not hard-and-fast rules, but can assist us to guide the process and articulate alternatives. They can lead us to think more deeply about our design rationale and prompt us to reflect. It helps to know “the rules” in order to wisely bend (or break) them at the right times!